

TARTU ÜLIKOOL  
HUMANITAARTEADUSTE JA KUNSTIDE VALDKOND  
EESTI JA ÜLDKEELETEADUSE INSTITUUT

Liisi Torga  
MITTE-PROJEKTIIVSED LAUSED EESTI KEELE SÕLTUVUSPUUDE PANGAS  
Bakalaureusetöö

Juhendaja: Kadri Muischnek

TARTU 2016

## Sisukord

Sissejuhatus.....	3
1. Sõltuvussüntaktiline lähenemine keeleteaduses ja arvutilingvistikas.....	4
1.1. Sõltuvus ja sõltuvussüntaksi põhitõed.....	4
1.2. Fraasistruktuuripuu ja sõltuvuspuu.....	5
1.3. Projektiivsus.....	5
1.4. Eesti keele sõltuvuspuude pank.....	6
2. Materjal.....	8
3. Analüüs.....	10
3.1. Verbiahela ristuvad kaared.....	10
3.1.1. Da-infinitiivne subjekt.....	10
3.1.2. Da-infiniivne objekt.....	12
3.1.3. Mitmesõnalise öeldise sõnajärg.....	13
3.1.4. Sõna mitmesõnalise öeldise osade vahel.....	13
3.1.5. Verb ja laiendid.....	14
3.1.6. Erandid.....	15
3.2. Sõnajärg.....	16
3.2.1. Koordinatsioon.....	17
3.2.2. Kõrvallause paigutus.....	17
3.2.3. Kaassõnafraas.....	19
3.2.4. Kvantorfraas.....	20
3.2.5. Võrdluskonstruksioon.....	20
3.2.6. Nimisõnafraasi sõnajärg.....	21
3.2.7. Lahkasendis fraasid.....	22
3.3. Ellips.....	22
3.4. Mittelauseliikmelised lauseelemendid.....	24
3.4.1. Üldlaiend.....	24
3.4.2. Kiil.....	25
3.5. Vead.....	25
3.5.1. Mitu juurt.....	26
3.5.2. Sulgudes element.....	26
3.5.3. Vale märgendamine.....	27
3.5.4. Vigased laused.....	27
Kokkuvõte.....	28
Kirjandus.....	30
Non-projective sentences in the Estonian Dependency Treebank.....	33

## Sissejuhatus

Puudepank ehk süntaktilise struktuuri suhtes märgendatud korpus kannab olulist rolli loomuliku keele automaattöötusel. Selle abil on võimalik arendada masintõlkesüsteeme ning testida nii keeletarkvarasid kui ka keeleteooriaid (Liin 2014). Eesti keele sõltuvuspuude pank koosneb morfoloogilisest, pindsüntaktilisest ja sõltuvussuhete kihist (Muischnek jt 2014b).

Käesolevas töös keskendutakse just sõltuvussuhetele. Suhteid kirjeldatakse kaarte abil, millega näidatakse, milline sõna allub millisele sõnale. Keeleteoreetilistes käsitlustes peetakse normiks, et kaared ei ristunud (nimetatakse **projektiivsuseks**), aga tunnistatakse, et kaarte ristumine (**mitte-projektiivsus**) on teatud keeleliste struktuuride esitamisel paratamatu (Nivre 2005). Arvutilingvistikas pigem välditakse selliseid ristumisi praktilistel põhjustel.

Bakalaureusetöö eesmärgiks on kirjeldada neid süntaktilisi struktuure, mis põhjustavad mitte-projektiivsust eesti keele sõltuvuspuude pangas. Esimeses peatükis antakse ülevaade sõltuvussüntaktilisest lähenemisest keeleteooriates ja arvutilingvistikas. See sisaldab endas sõltuvusteooria põhimõistete lahtiseletamist, sõltuvuspuu ja fraasistruktuuripuu võrdlust, projektiivsuse lahtimõtestamist ning ülevaadet eesti keele sõltuvuspuude pangast ja seal esinevatest problemaatilistest keelenditest. Teises peatükis kirjeldatakse materjali ja töö käiku. Kolmas peatükk kirjeldab neid süntaktilisi struktuure, mis põhjustavad mitte-projektiivsust eesti keele sõltuvuspuude pangas. Neljandas peatükis esitatakse kokkuvõtlikult kogu töö ülevaade.

# 1. Sõltuvussüntaktiline lähenemine keeleteaduses ja arvutilingvistikas

## 1.1. Sõltuvus ja sõltuvussüntaksi põhitõed

Sõltuvusanalüüsi keskseks põhimõtteks on see, et lause süntaktiline struktuur koosneb binaarsetest ebasümmeetrilistest suhetest. Binaarne ebasümmeetriline suhe on kahe tekstisõna vaheline suhe, kus üks suhte osaline on ülemus (*head*) ja teine on alluv (*dependent*). (Nivre 2005)

Sõltuvuspuudes sõltuvussuhete kirjeldamiseks kasutatakse mõisteid tipp, kaar ja juur. Sõltuvuspuudes on tipuks (*node*) sõnavorm lauses. Kahte tippu ühendab kaar (*arc*), millel on informatsioon süntaktilise suhte kohta. Juur või juurtipp (*root, top node*) on selline tipp, millel puudub ülemus. (Melčuk 1988: 23)

Nivre (2005) kirjeldab sõltuvussuhteid graafiteooriat kasutades, kus eeldatakse, et graaf peab olema sidus (st iga tipp peab olema seotud vähemalt ühe teise tipuga), lisaks sellele eeldatakse, et tipul on kõige enam üks ülemus (ühe pea kitsendus ehk *single-head constraint*) ning et graafi tippude vahel ei tohi olla tsüklit (mitte-tsüklilisuse kitsendus ehk *acyclicity constraint*). Nendest tingimustest järeldub see, et esineb selline tipp, millel puudub ülemus. Sellist tippu nimetatakse juurtipuks.

Melčuk (1988: 21–22) kirjeldab, kuidas süntaktilised suhted peavad olema antisümmeetrilised, mis tähendab seda, et kui sõnapaaris olevad sõnad ümber pöörata, siis nende semantiline tähendus algsest muutub. Suhted peavad olema ka antirefleksiivsed, mis tähendab, et ükski sõna ei saa olla iseenda alluv. Kolmandaks tingimuseks on see, et suhted peavad olema antitransiivsed, mis tähendab, et sõna alluva alluv ei saa olla selle sõna otsene alluv.

## 1.2. Fraasistruktuuripuu ja sõltuvuspuu

Kuigi käesolevas töös kasutatakse sõltuvuspuid lausete süntaktilise struktuuri kirjeldamiseks, siis on olemas ka teine, vastandlik viis, kuidas struktuuri kirjeldada – fraasistruktuuripuud. Järgnevalt tuuakse välja nende viiside peamised erinevused.

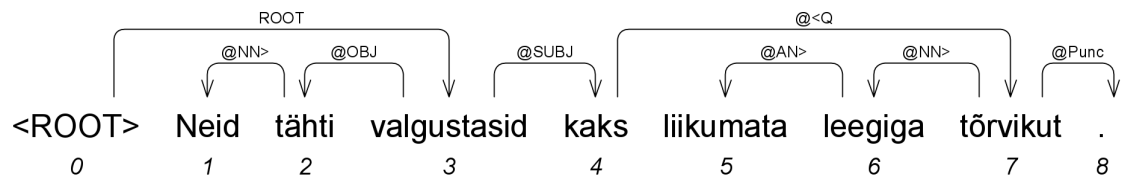
Sõltuvuspuid eristab fraasistruktuuripuudest see, kuidas süntaktiline struktuur neis avaldub. Kui sõltuvuspuu avaldab lause struktuuri binaarsete sõltuvussuhete kaudu, siis fraasistruktuuripuu kujutab endas hierarhilist struktuuri, kus sõnavormid või fraasid moodustavad kokkukuuluvaid üksusi, mis omakorda moodustavad suuremaid kokkukuuluvaid üksusi.

Kui sõltuvuspuus on rõhk pandud binaarsetele sõltuvussuhetele, siis fraasistruktuuripuu rõhutab üksuste klassikuuluvust. Neid kategoriseeritakse abstraktsete tippude abil (*NP* = *noun phrase* ehk nimisõnafraas, *VP* = *verb phrase* ehk verbifraas jne). Suurem osa tippudest on fraasistruktuuripuu abstraktsed. Seevastu sõltuvuspuudes klassikuuluvust ei kirjeldata ega kasutata abstraktseid tippe. (Melčuk 1988: 13–15)

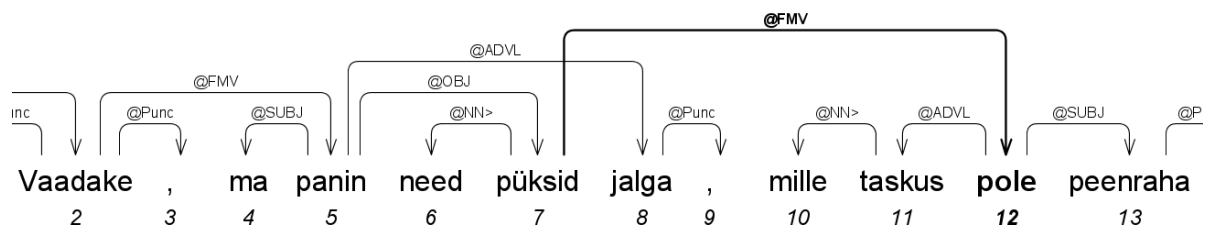
Melčuk (1988: 16–17) toob välja, et 18-sõnalise lause kirjeldamiseks läheb fraasistruktuuripuu vaja 61 tippu ja sõltuvuspuus ainult 18. Samuti on fraasistruktuuripuu oluline tippude järjekord, aga sõltuvuspuus ei ole. See tähendab, et sõltuvuspuu on tunduvalt ökonoomsem kirjeldamaks lausete struktuuri.

## 1.3. Projektiivsus

Lause on projektiivne, kui ükski kaar omavahel ei ristunud ega kata juurtippu. Enamik loomuliku keele lausetest on projektiivsed, kuid esineb ka mitte-projektiivsust. Mitte-projektiivsed laused on Melčuki sõnul alati mingil kombel markeeritud: ristuvaid kaari põhjustavad stiil, rõhk, erilised süntaktilised elemendid, nagu nt kliitikud. (Melčuk 1988: 35–36)



Joonis 1. Projektiivne lause, kus kaared ei ristu (ilukirjandusklass; EKSP: ilu\_kivirahk.inforem)



Joonis 2. Mitte-projektiivne lause, kus kaared ristuvad (ilukirjandusklass; EKSP: ilu\_sauter\_10000.inforem)

Nivre (2005) kirjeldab, et sõltuvuse keeleteoreetilistes käsitlustes on projektiivsus normiks, kuid peetakse paratamatuks ka mitte-projektiivsuse olemasolu teatud keelekonstruktidest. Praktistes käsitlustes (nt arvutilingvistikas) pigem välditakse mitte-projektiivsust praktilistel põhjustel.

#### 1.4. Eesti keele sõltuvuspuude pank

Puudepankadeks nimetatakse selliseid korpuseid, kus kõik laused on süntaktiliselt märgendatud (Jurafsky 2009). Eesti keele sõltuvuspuude pank kujutab endas korpust, kus on umbes 400 000 sõna. Puudepanga märgendus koosneb kolmest kihist: morfoloogilisest kihist, kus esitatakse iga sõna grammatiline info, süntaktilisest kihist, kus esitatakse iga sõna süntaktiline info ja sõltuvussuhete kihist, kus kirjeldatakse sõnadevahelisi ülemus-alluvus suhteid. (Muischnek jt 2014b)

Lausete analüüsimiseks kasutati eesti keele kitsenduste grammatika sõltuvussüntaktilist analüsaatorit (Muischnek jt 2014a). Saadud tulemusi parandasid ja kontrollisid käsitsi kaks märgendajat. Erinevuste tekkimisel kontrollis lauseid ka kolmas märgendaja, kelle

ülesandeks oli erimeelsuste lahendamine. (Muischnek jt 2014b) Eesti keele sõltuvuspuude pank on märgendatud kasutades eesti keele kitsenduste grammatika (Müürisep 2000) süntaktilisi märgendeid ning igale tekstisõnale on lisatud info tema sõltuvusstruktuurilise ülemuse kohta (Muischnek jt 2014a).

Eesti keele sõnajärg on vaba, peamiselt määratletud infostruktuurist. Verb asub lauses teisel kohal. Mitmesõnalise öeldise puhul asub teisel kohal öeldise finiidne komponent ja perifrastiline verbi osis ja põhiverb paiknevad lause lõpus, verbi seotud ja vabade laiendite järel. (Lindström (ilmumas))

Puudepankade märgendusel on esile kerkinud teatud keelendid, mis lausetes mitteprojektiivsust tekitavad. Nendeks on koordinatsioon, mitmesõnalised predikaadid, kaassõnafraasid ja kvantorfraasid. (McDonald *et al* 2013)

Tuginedes sõltuvussüntaktiliselt analüüsitud korpuse märgendusjuhendile, on koordinatsioon märgendatud nii, et iga järgnev koordineeritud üksus allub eelnevale. Sidesõna märgendatakse nii, et see allub talle järgnevale koordineeritud üksusele. Ka koordineeritud osalused märgendatakse nii, et koordineeritud osalause predikaadile allub talle järgneva koordineeritud osalause predikaat. (Märgendusjuhend)

Perifrastilised verbid märgendatakse nii, et ülemuseks on infiniitne verbivorm. Rohkem kui kahest elemendist koosnevad perifrastilised verbid on märgendatud ahelana. Verbi ja ta laiendite puhul aga on ülemuseks finiidne verb. Verbi liitvormis on kõrgeimaks ülemuseks infiniit, mis tingib lauses olevad seotud laiendid. (Märgendusjuhend)

Kvantorfraasi puhul on ülemuseks kvantor ja talle allub ühendisse kuuluv noomen. Sama on ka kaassõnafraasiga, kus ülemuseks on kaassõna ja talle allub ühendisse kuuluv noomen. (Märgendusjuhend)

## 2. Materjal

Käesolevas töös analüüsitud laused on pärit eesti keele sõltuvuspuude pangast. Puudepangas on 30 299 lauset (~400 000 sõna) ajakirjandusest, ilukirjandusest, populaarteaduslikest ja teaduslikest tekstidest. Nimekiri failidest ja nende tekstiklassidest asub puudepanga kodulehel (vt viide EKSP). Analüüsiti 15 464 lauset, mis moodustab enam-vähem poole korpusest. Analüüsimiseks kasutati programmi MaltEval Tree Viewer 1.0.1. Tegemist on programmiga, mis võimaldab graafiliselt kujutada sõltuvuspuid ning tõsta esile huvipakkuvat informatsiooni (Nilsson jt 2008). Käesolevas töös sorteeriti automaatselt välja laused, mis sisaldasid mitte-projektiivsust ning lauses olevad ristuvad kaared tõsteti esile.

*Tabel 1. Lausete analüüs ja mitte-projektiivsete lausete hulk*

	<b>Lauseid korpuses</b>	<b>Laused, mida analüüsiti</b>	<b>Analüüsitud lausetest mitte-projektiivsed</b>
<b>Ilukirjandusteksid</b>	5813	5033	159
<b>Teadustekstid</b>	5790	2960	237
<b>Ajakirjandustekstid</b>	18 696	7471	343
<b>Kokku</b>	30299	15464	739

Mitte-projektiivsust esines 739 lauses 15 464 lausest, st 100 lausest 4.78 lauses esines üks või mitu kaarte ristumist (vt Tabel 1). Kõige rohkem mitte-projektiivsust esines teadustekstides. Üks olulisemaid põhjuseid on see, et nendes tekstides on väga palju selliseid elemente, mida keeleliselt analüüsida ei saa ning mis sel põhjusel jäeti esialgu käsitsi kontrollitud märgendamisel ülejäänud lausega ühendamata. Teisendades neid puudepanka ConLL-formaati, et oleks võimalik neid programmiga MaltEval visualiseerida, ühendati sellised elemendid ülejäänud lausega mõnevõrra ebajärjekindlalt. Erinevatel failidel rakendati erinevaid põhimõtteid. (Muischnek 2016)

Lauseid analüüsides leiti viis selgelt eristuvat mitte-projektiivsust põhjustavat gruppi, mis omakorda jagunesid alagruppideks (vt Tabel 2). Nendest järgmises peatükis lähemalt.



Tabel 2. Ristumiste arv liigiti

		Ristumiste arv	Ristumiste arv 100 lause kohta
<b>Verbiahelad</b>	<b>Da-infinittiivne subjekt</b>	153	0.99
	<b>Da-infinittiivne objekt</b>	74	0.48
	<b>Mitmesõnalise öeldise sõnajärg</b>	16	0.10
	<b>Sõna mitmesõnalise öeldise osade vahel</b>	21	0.14
	<b>Verb ja laiendid</b>	27	0.17
	<b>Erandid</b>	5	0.03
	<b>Verbiahelad kokku</b>	296	1.91
<b>Sõnajärg</b>	<b>Koordinatsioon</b>	4	0.03
	<b>Kõrvallause paigutus</b>	22	0.14
	<b>Kaassõnafraas</b>	37	0.24
	<b>Kvantorfraas</b>	10	0.06
	<b>Võrdluskonstruktsioon</b>	15	0.10
	<b>Nimisõnafraasi sõnajärg</b>	11	0.07
	<b>Lahkasendis fraasid</b>	15	0.10
	<b>Sõnajärg kokku</b>	114	0.74
<b>Ellips</b>	<b>Lühendus- ja kontekstiellipsid</b>	75	0.48
<b>Mittelause-liikmelised lauseelemendid</b>	<b>Üldlaiend</b>	11	0.07
	<b>Kiil</b>	6	0.04
	<b>Mittelauseliikmelised lauseelemendid kokku</b>	17	0.11
<b>Vead</b>	<b>Mitu juurt</b>	22	0.14
	<b>Sulgudes element</b>	100	0.65
	<b>Vale märgendamine</b>	124	0.80
	<b>Vigased laused</b>	8	0.05
	<b>Vead kokku</b>	254	1.64
<b>Ristuvad kaared kokku</b>		756	4.89

### 3. Analüüs

#### 3.1. Verbiahela ristuvad kaared

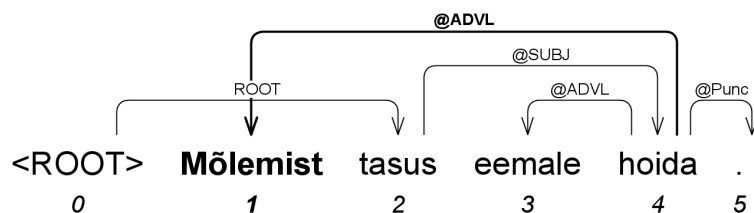
Kõige suuremaks mitte-projektiivsust põhjustavaks grupiks (kui märgendamise vead välja jätta) on verbiahelad. Kokku esines 294 verbiahelatest tulenevat kaarte ristumist, st 100 lause kohta esines mitte-projektiivsust 1.90 korral. Verbiahela all peetakse silmas mitmest verbivormist koosnevat süntaktilist struktuuri, mis võivad olla nii ahelverbid kui ka verbist ja tema argumendist (subjekt, objekt, adverbiaal) koosnevad ühendid. Ahelverb on öeldisena toimida võiv finiiitse ja infiniitse verbi ühend, mille sisu kannab infiniitne verb (EKG II: 19). Verbiahelad jagunevad omakorda viieks eraldiseisvaks grupiks.

##### 3.1.1. *Da-infinitiivne subjekt*

Da-infinitiivitarindi ülesandeks on laiendada predikaati, sealjuures on tema funktsiooniks olla kas subjekt, objekt, adverbiaal, atribuut või predikatiiv. Da-infinitiiv subjektina esineb peamiselt kogejalauses. (EKG II: 237) Kokku esines puudepangas 153 da-infinitiivse subjekti põhjustatud kaarte ristumist, st 100 lause kohta esines mitte-projektiivsust 0.99 korral.

Puudepangas esineb lauseid, kus finiiitset verbi laiendab da-infinitiivne subjekt, millel on objekt või muu seotud laiend teisel pool osalause juurtippu. Mitte-projektiivsust esineb sellistes lauses, kus da-infinitiivne subjekt laiendab vajadust/kohustust väljendavaid predikaate (nt *vaja/tarvis* + *olema*, *tulema*, *pruukima*, *tasuma* jne) ja suutelisusele viitavaid predikaate (nt *õnnestuma*) (EKG II: 238–239).

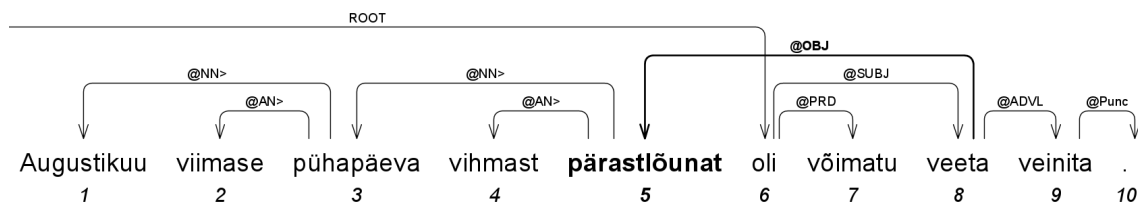
Joonisel 3 on vajadust väljendav finiiitne verbivorm *tasus*, millele allub da-infinitiivne subjekt *hoida*, millele omakorda allub teisel pool juurt olev adverbiaal *mõlemist*, mis ristub finiiitse verbivormiga *tasus*.



Joonis 3. Näitelause ilukirjandusklassist (EKSP: *ilu\_valton\_kana\_10000.inforem*)

Esineb ka lauseid, kus finiiitse verbi juurde kuulub da-infinitiivne subjekt, millel on objekt või muu seotud laiend teisel pool osalause juurtippu, kusjuures finiiitse verbi juurde kuulub predikatiiv. Mitte-projektiivsust esineb sellistes lausetes, kus da-infinitiivne subjekt laiendab kas emotsionaalset/füsioloogilist seisundit väljendavaid predikaate (nt *olema* + *kole*, *ilus*, *raske* jne) või sobivust väljendavaid predikaate (nt *olema* + *võimatu*, *sobilik* jne) (EKG II: 237–238).

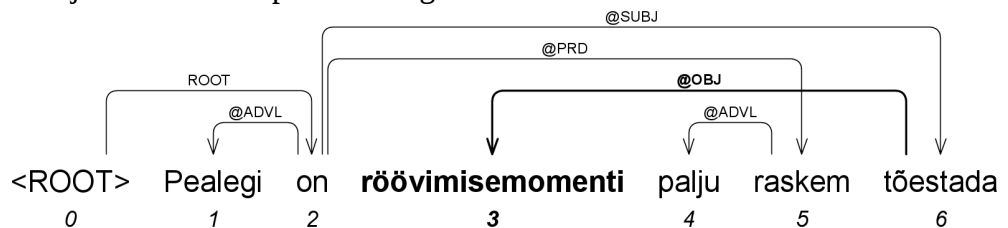
Joonisel 4 on finiiitne verbivorm *oli*, millele allub sobivust väljendav predikatiiv *võimatu*, aga ka da-infinitiivne subjekt *veeta*, millele omakorda allub teisel pool juurt olev objekt *pärastlõunat*, mis ristub finiiitse verbivormiga *oli*.



Joonis 4. Näitelause ajakirjandusklassist (EKSP: *aja\_pm\_1998\_09\_26e.inforem*)

Esineb grupp lauseid, kus on finiiitne verb, mida laiendab predikatiiv, aga ka da-infinitiivne subjekt, millele kuuluv objekt või muu laiend ristub samal pool juurt oleva predikatiiviga. See tähendab, et da-infinitiivsele subjektile kuuluv objekt või muu laiend asetseb finiiitse verbi ja talle kuuluva predikatiivi vahel. Vahel esineb ka topeltristumist, mis on põhjustatud sellest, et lisaks predikatiiviga ristumisele, ristub da-infinitiivse subjekti objekt või muu laiend ka finiiitse verbi külge kuuluva muu laiendiga.

Joonisel 5 on finiidne verbivorm *on*, millele allub predikatiiv *raskem*, aga ka da-infiniivne subjekt *tõestada*, millele omakorda allub objekt *rõõvimisemomenti*, mis ristub finiidse verbi juurde kuuluva predikatiiviga.

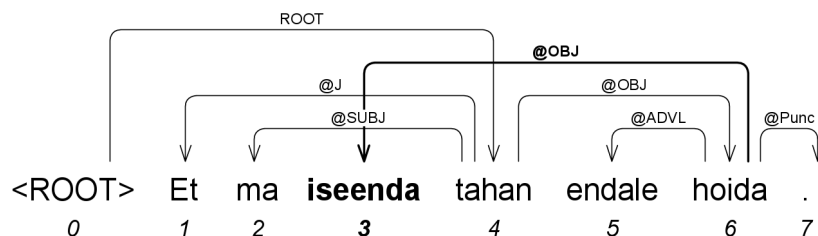


Joonis 5. Näitelause ajakirjandusklassist (EKSP: aja\_ee\_2001\_10.tasak.inforem)

### 3.1.2. Da-infiniivne objekt

Da-infiniitvne objekt põhjustab mitte-projektiivsust sellistes lausetes, kus ta laiendab soovi, kavatsust või üritamist väljendavaid predikaate (*soovima, tahtma, kavatsema* jne) või julgemist, oskamist, suutmist väljendavaid predikaate (*oskama, mõistma, jõudma* jne) (EKG II: 240). Da-infiniitvne objekt võib laiendada ka selliseid predikaate nagu *käskima, aitama, soovitama* jne (EKG: 241). Kokku esines puudepangas 74 da-infiniitvse objekti põhjustatud kaarte ristumist, st 100 lause kohta esines mitte-projektiivsust 0.48 korral.

Ristumine toimub sellisel juhul, kui da-infiniitvsele objektile alluv objekt või muu laiend asub teisel pool juurtippu, mis põhjustab ristumise. Joonisel 6 on finiidne verbivorm *tahan*, millele allub da-infiniitvne objekt *hoida*, millele omakorda allub objekt *iseenda*, mis ristub finiidse verbi ja talle kuuluvate subjekti ja konjunktsiooniga.

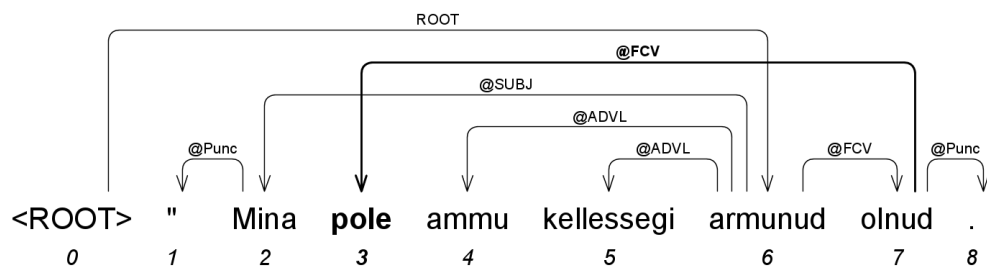


Joonis 6. Näitelause ilukirjandusklassist (EKSP: ilu\_sauter\_10000.inforem)

### 3.1.3. Mitmesõnalise öeldise sõnajärg

Lausetes esinevad sellised mitmesõnalised öeldised (*sai tegema hakatud, oleks sulama hakanud, pole armunud olnud*), kus ühendi põhisisu kandev infiniitne verbivorm on tõstetud ettepoole. Seda põhjustab infostruktuur ja rõhutamine. Kokku esines puudepangas 16 mitmesõnalise öeldise sõnajärjest põhjustatud kaarte ristumist, st 100 lause kohta esines mitte-projektiivsust 0.10 korral.

Joonisel 7 verbifraasi *pole armunud olnud* neutraalne sõnajärg oleks *pole olnud armunud*. Selliseid kolmekomponendilisi öeldisi märgendatakse ahelana. Joonisel 7 on kõrgeimaks ülemuseks *armunud*, sellele allub *olnud*, millele omakorda allub *pole*. Sisu kandvale verbile kuuluvad ka subjekt ja teised verbile kuuluvad laiendid. Kuna verbiahelas muutub sõnajärg, aga alluvussuhted jäävad samaks, siis tekibki ristumine.

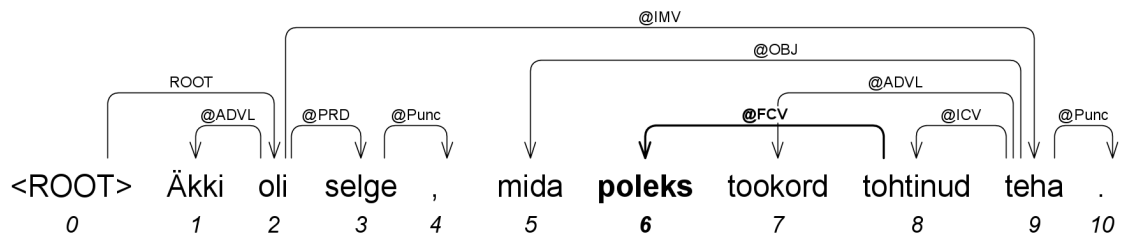


Joonis 7. Näitelause ilukirjandusklassist (EKSP: *ilu\_maailm.inforem*)

### 3.1.4. Sõna mitmesõnalise öeldise osade vahel

Ka siin on tegemist mitmesõnaliste öeldistega (*oled hakanud pidama, poleks tohtinud teha, peavad olema registreeritud*), aga siin olevates lausetes on neil neutraalne sõnajärg. Probleemiks on see, et kolmekomponendilise öeldise kõrgeimale ülemusele kuuluv alluv (see võib olla nii subjekt, objekt kui ka adverbiaal) jääb öeldise esimese ja teise komponendi vahele. Kokku esines puudepangas 21 mitmesõnalise öeldise vahele jääva sõna poolt põhjustatud kaarte ristumist, st 100 lause kohta esines mitteprojektiivsust 0.14 korral.

Joonisel 8 on kõrgeimaks ülemuseks *teha*, sellele allub *tohtinud*, millele omakorda allub *poleks*. Ristumise põhjustab kõrgeimale ülemusele kuuluv adverbiaal *tookord*, mis asub *poleks* ja *tohtinud* vahel.



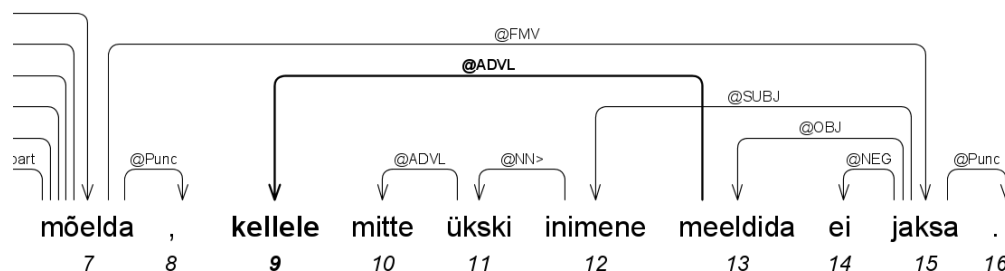
Joonis 8. Näitelause ilukirjandusklassist (EKSP: *ilu\_maailm.inforem*)

### 3.1.5. Verb ja laiendid

Siia gruppi kuuluvad need verbistruktuurid, mis koosnevad kahest verbist, kusjuures teine verb on kas objekt või mõni muu seotud laiend. Verbi ja ta laiendi puhul märgendatakse finiidne verb ülemuseks ja talle allub objekt või muu seotud laiend (Märgendusjuhend). Kokku esines puudepangas 27 verbi ja ta laiendite põhjustatud kaarte ristumist, st 100 lause kohta esines mitte-projektiivsust 0.17 korral.

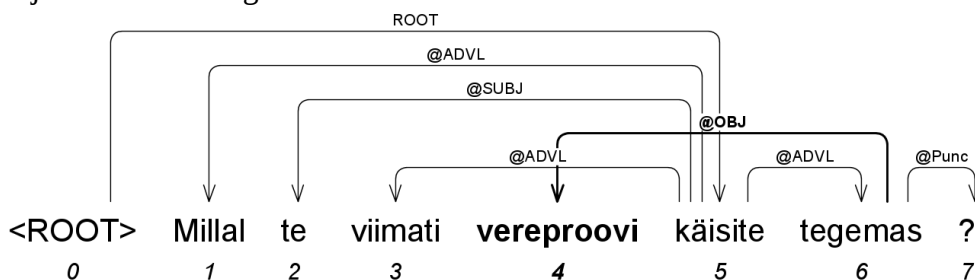
Esineb lauseid, kus öeldisverbile allub da-infinitiivne objekt. Nii öeldisverbile kui talle alluvale da-infinitiivsele objektile kuuluvad omakorda alluvad, mis lauses asetsevad vaheldumisi. Seda gruppi iseloomustab ka öeldisverbi rõhutamine, mille tõttu paigutub öeldisverb paremale.

Joonisel 9 on osalause verbivormiks *jaksa*, mille objektiks on *meeldida*. Ristumine tekib sellest, et mõlemale verbile alluvad laiendid (*kellele*, *mitte ükski inimene*) paiknevad osalauses vaheldumisi. Da-infinitiivsele objektile *meeldida* alluv *kellele* kannab ka sidendi funktsiooni, mille tõttu asub ta osalause alguses ning tekitab ristumise.



Joonis 9. Näitelause ilukirjandusklassist (EKSP: ilu\_sauter\_10000.inforem)

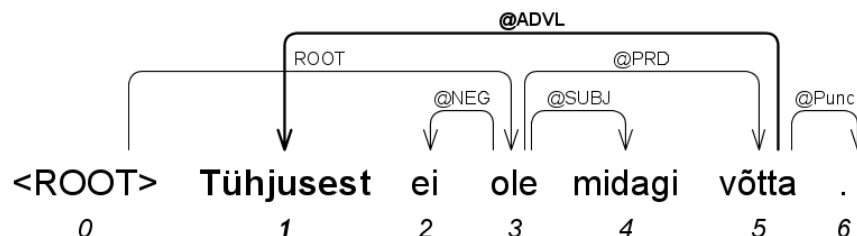
Esineb ka selliseid verbiahelaid, kus finiiitset verbi laiendab ma-infinitiiv (nii ma-, mas-, mast- kui ka mata-tarindid). Ma-infinitiiv käitub lauses adverbialina. Puudepangas esineb lauseid, kus ma-tarind laiendab nt liikumisverbe *minema*, *tulema*, *käima*. (EKG II 251–252) Joonisel 10 on finiiitseks verbivormiks *käisite*, millele allub adverbial *tegemas*, millele omakorda allub objekt *vereproovi*, mis põhjustab ristumise finiiitse verbi ja tema laienditega.



Joonis 10. Näitelause ilukirjandusklassist (EKSP: ilu\_maailm.inforem)

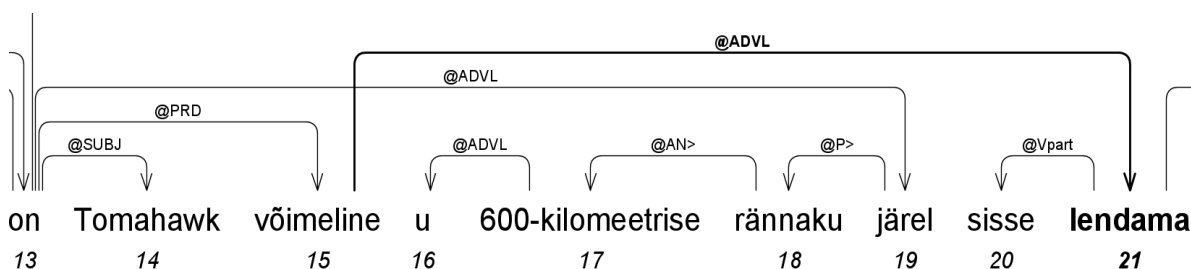
### 3.1.6. Erandid

Siia gruppi koonduvad need verbiahela näited, mis on erandlikud, kuid siiski olulised. Erandite põhjustatud kaarte ristumisi esines kokku 5, st 100 lause kohta esines mitteprojektiivsust 0.03 korral. Nii nagu lausetes esineb da-infiniitvset subjekti ja objekti, esineb ka da-infiniitvset predikatiivi, mis laiendab finiiitset verbi *olema*. Joonisel 11 on finiiitne verbivorm *ole*, millele allub da-infiniitvne predikatiiv *võtta*, millele omakorda allub adverbial *tühjusest*, mis ristub finiiitse verbiga.



Joonis 11. Näitelause ilukirjandusklassist (EKSP: ilu\_remsu\_siiditeel.inforem)

Esineb ka lauseid, kus verb laiendab predikatiivi. Peamiselt laiendas verb adjektiivi *võimeline*. Joonisel 12 on finiitne verbivorm *on*, mida laiendab predikatiiv *võimeline*, mida omakorda laiendab adverbiaal *lendama*. Ristumine tekib sellest, et finiitse verbivormi külge kuuluvad laiendid asuvad predikatiivi ja teda laiendava adverbiaali vahel.



Joonis 12. Näitelause ajakirjandusklassist (EKSP: aja\_luup\_2000\_09.inforem)

### 3.2. Sõnajärg

Kuigi sõnajärg on keskseks probleemiks pea kõikides gruppides, siis siin on välja toodud need sõnajärjeprobleemid, mis teistesse gruppidesse ei sobi. Kokku esines puudepangas 114 sõnajärjest põhjustatud kaarte ristumist, st 100 lause kohta esines mitte-projektiivsust 0.74 korral. Sõnajärjest tulenevad kaarte ristumised jagunevad seitsmeks eraldiseisvaks grupiks.

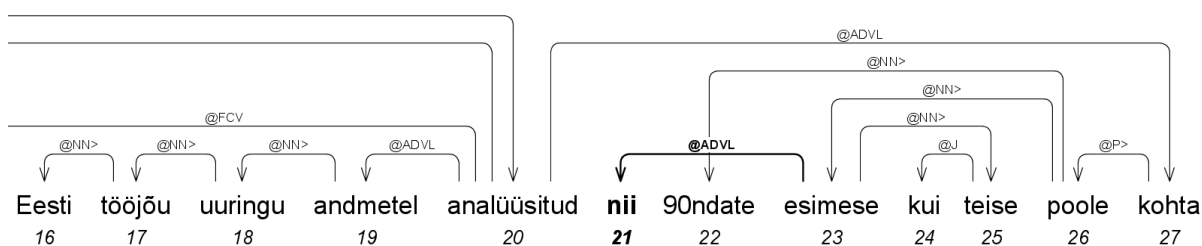


### 3.2.1. Koordinatsioon

Rinnastuse ehk koordinatsiooni ülesandeks on siduda süntaktiliselt samaväärseid moodustajaid. Rinnastatud rindliikmed moodustavad kõrvuti asetseva rindtarindi. Moodustajad võivad olla nii mittelauselised (koondlause) kui ka rinnastatud osalaused (rindlause). (EKG II: 213) Kokku esines puudepangas 4 koordinatsiooni põhjustatud kaarte ristumist, st 100 lause kohta esines mitte-projektiivsust 0.03 korral.

Puudepangas märgendatakse koordinatsiooni nii, et esimene koordineeritud element on põhi ja järgmine koordineeritud element sõltub eelnevast. Sidendi ülemuseks on talle järgnev koordineeritud element. (Märgendusjuhend) Ristumine tekib sel juhul, kui koordineeritud elemendil on täiendid ning sidend asub täiendite ja koordineeritud elemendi vahel või sidendi ja koordineeritud elemendi vahele jääb mõni teine sõna, mis ei allu kummalegi.

Joonisel 13 koordineeritakse täiendid *esimese* ja *teise*, ristumise põhjustab sidendi *nii* ja tema ülemuseks oleva *esimese* vahel olev täiend *90ndate*, mis ei allu kummalegi.



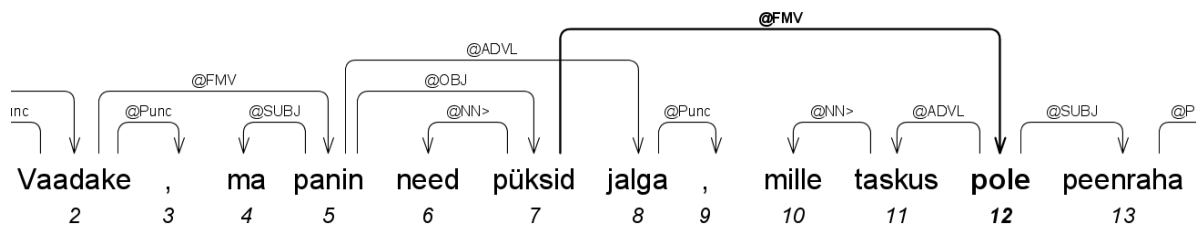
Joonis 13. Näitelause teadusklassist (EKSP: tea\_toohoive\_10000.inforem)

### 3.2.2. Kõrvallause paigutus

Põimlause on liitlause, mis koosneb põhilausest ja kõrvallausest. Pealause on selline põhilause, millel on ainult alistavad tunnused. Pealauset ja kõrvallauset seovad sidend ja korrelaat. (EKG II: 281) Relatiivlause on selline kõrvallause, mida seob pealausega

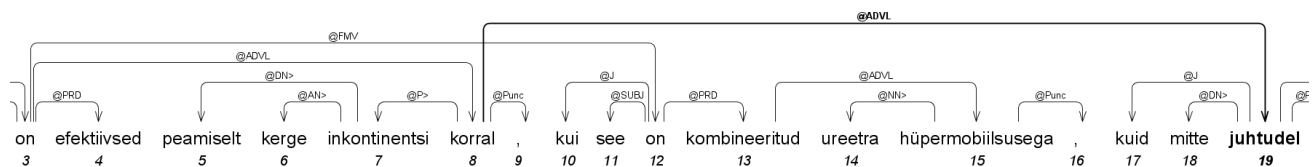
pronoomen või adverb (EKG II: 311). Kokku esines puudepangas 22 kõrvallause paigutusest põhjustatud kaarte ristumist, st 100 lause kohta esines mitte-projektiivsust 0.14 korral.

Puudepangas märgendatakse atributiivseid relatiivlauseid nii, et kõrvallause predikaat allub pealause viimasel positsioonil olevale noomenile, mida kõrvallause laiendab (Märgendusjuhend). Ristuvad kaared tekivad sel juhul, kui sõna või fraas, mida relatiivlause laiendab, ei asu pealause viimasel kohal. Joonisel 14 on kõrvallauseks *mille taskus pole peenraha*, kus sidendiks on *mille* ning kõrvallause laiendab objekti *püksid*. Ristumine tekib sellest, et pärast objekti *püksid* on veel adverbiaal *jalga*, mis kuulub pealause predikaadile.



Joonis 14. Näitelause ilukirjandusklassist (EKSP: ilu\_sauter\_10000.inforem)

Esineb ka selliseid lauseid, kus on rinnastatud mingid lauseliikmed ning nende vahele on paigutatud osalause. Joonisel 15 on rinnastatud adverbiaalid *korral* ja *juhtudel* ning nende vahele on paigutatud osalause *kui see on kombineeritud ureetra hüpermobiilsusega*, mis riputatakse pealause predikaadi külge.

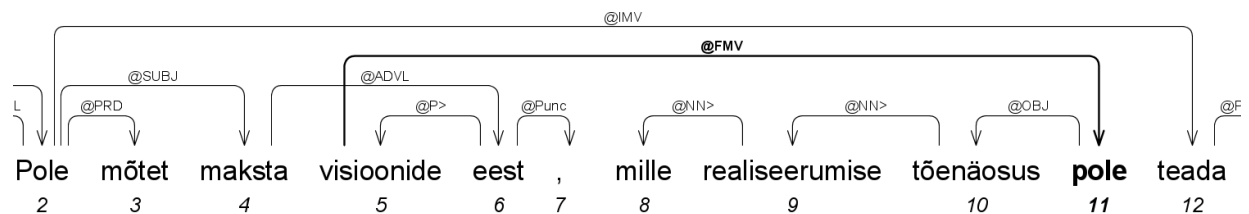


Joonis 15. Näitelause teadusklassist (EKSP: tea\_eesti\_arst\_2004.inforem)

### 3.2.3. Kaassõnafraas

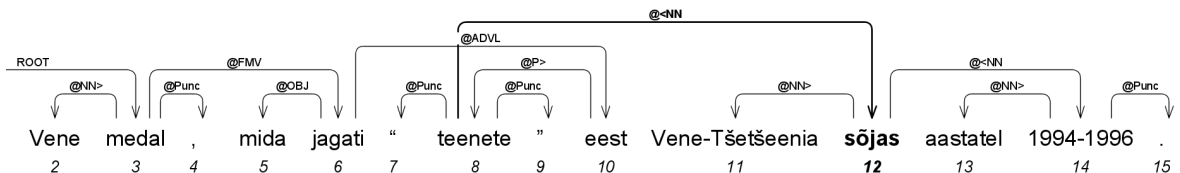
Kaassõnafraasi moodustavad kaassõna ja substantiiv või substantiivifraas, kus põhjaks on kaassõna, mis tingib substantiivi käände. Kaassõnafraasi moodustavateks kaassõnadeks võivad olla nt *järel*, *peal*, *kujul*, *puhul*, *peale*, *eest* jne. (EKG II: 137–139) Puudepangas märgendatakse kaassõnafraase nii, et kaassõnale allub substantiiv või substantiivifraas (Märgendusjuhend). Kokku esines puudepangas 37 kaassõnafraasi põhjustatud kaarte ristumist, st 100 lause kohta esines mitte-projektiivsust 0.24 korral. Lausetes esines peamiselt kaht tüüpi kaassõnafraasi kaarte ristumisi.

Esineb lauseid, kus kaassõnafraas asub osalausepiiril. Ristumine tekib seetõttu, et substantiivi või substantiivifraasi ja seda laiendava osalause vahele jääb kaassõna. Joonisel 16 on kaassõnafraasiks *visioonide eest*, kus kaassõnaks on *eest* ja sellele allub substantiiv *visioonide*, mille külge riputatakse talle kuuluv osalause.



Joonis 16. Näitelause ajakirjandusklassist (EKSP: aja\_luup\_2000\_09.inforem)

Esineb ka selliseid lauseid, kus kaassõnafraasi substantiivi ja temale kuuluvate atribuutide vahele jääb kaassõna. Joonisel 17 on kaassõnafraasiks *teenete eest*, kus kaassõnaks on *eest* ja talle allub substantiiv *teenete*, millele omakorda kuuluvad atribuudid *Vene-Tšetšeenia sõjas aastatel 1994-1996*, mis põhjustavad ristumise kaassõnaga.



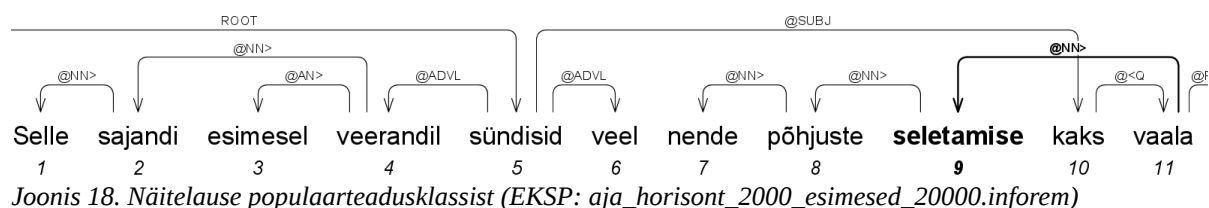
Joonis 17. Näitelause populaarteadusklassist (EKSP: aja\_horisont\_2000\_esimesed\_20000.inforem)

Kui kaassõnafraasid märgendada nii, et põhjaks oleks hoopiski substantiiv ja talle alluks kaassõna, siis üldjuhul jääksid ristumised ära.

### 3.2.4. Kvantorfraas

Kvantorfraasi moodustavad kvantor ja substantiiv, kus kvantor on põhjaks ning see põhjustab substantiivi partitiivse käände. Kvantorfraasid väljendavad mõõtu, hulka, määra. (EKG II: 140) Kokku esines puudepangas 10 kvantorfraasi põhjustatud kaarte ristumist, st 100 lause kohta esines mitte-projektiivsust 0.06 korral.

Ka puudepangas märgendatakse kvantorfraase nii, et laiendiks olev substantiiv allub kvantorile (Märgendusjuhend). Ristuvaid kaari põhjustab peamiselt see, et substantiivi ja tema atribuudi või atribuutide vahele jääb kvantor. Joonisel 18 on kvantorfraas *kaks vaala*, kus kvantoriks on *kaks* ja teda laiendab substantiiv *vaala*, mille juurde kuuluvad atribuudid *nende põhjuste seletamise*, mis tekitavad ristumise. Ka kvantorfraasi kohta saab öelda, et kui põhjaks oleks substantiiv ja talle alluks kvantor, siis ristumist ei tekiks.

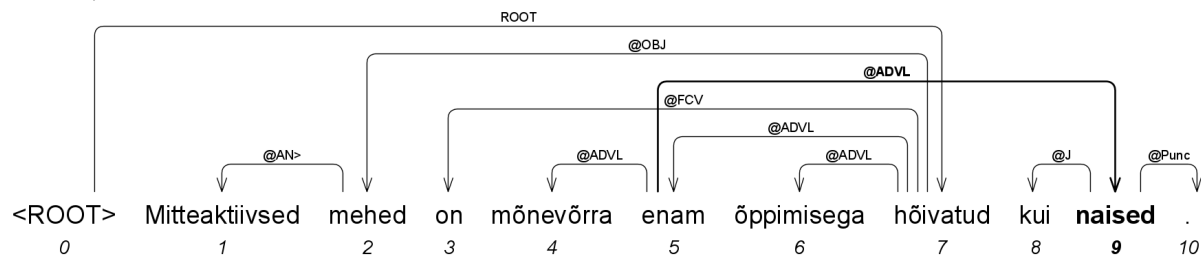


### 3.2.5. Võrdluskonstruktsioon

Komparatiivtarindi puhul võrreldakse objektide mingi omaduse astme tugevust. Võrdlusalust fraasi ja adjektiivi seob konjunktsioon *kui*. Võrdluse edasiandmiseks kasutatakse mõnikord adjektiivi asemel ka adverbide, nagu nt *rohkem*, *vähem*, *enam*. (EKG II: 132–133) Kokku esines puudepangas 15 võrdluskonstruktsiooni põhjustatud kaarte ristumist, st 100 lause kohta esines mitte-projektiivsust 0.10 korral.

Puudepangas märgendatakse võrdluskonstruktsioone nii, et võrdlust edasi andva sõna (adjektiiv või adverb) külge riputatakse kompareeritav objekt pärast konjunktsiooni. Ristumine tekib sellisel juhul, kui komparatiivvormis adjektiiv või adverb ja konjunktsioon pole lauses kõrvuti. Joonisel 19 on võrdlust edasi andvaks adverbiks

*enam*, millele allub kompareeritav objekt *naised*. Adverbi *enam* ja konjunktsiooni *kui* vahele jääb verbi liitvormi osa *õppimisega hõivatud*, mille tõttu kaared ristuvad. Kui lause sõnajärg oleks *Mitteaktiivsed mehed on õppimisega hõivatud mõnevõrra enam kui naised*, siis ristumist ei toimuks.

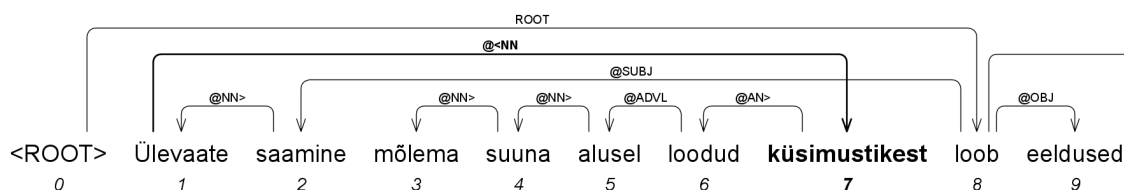


Joonis 19. Näitelause teadusklassist (EKSP: tea\_toohoive\_10000.inforem)

### 3.2.6. Nimisõnafraasi sõnajärg

Nimisõnafraasiks ehk substantiivifraasiks nimetatakse fraasi, mille põhjaks on substantiiv ning talle võivad kuuluda täiendid ehk atribuudid või lisandid ehk apositsioonid. (EKG II: 114) Kokku esines puudepangas 11 nimisõnafraasi sõnajärjest põhjustatud kaarte ristumist, st 100 lause kohta esines mitte-projektiivsust 0.07 korral.

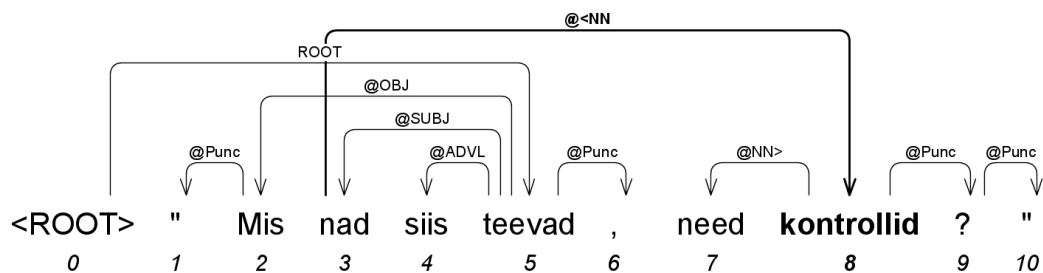
Puudepangas tekivad ristumised fraasi sõnajärjest. Teatud juhtudel esineb ka nominalisatsiooni, mis tingib nimisõnafraasi sõnajärje. Nominalisatsioon tuletatakse finiidest verbivormist (EKG II: 269). Joonisel 20 moodustab nominaaltarindi *ülevaate saamine*, mis on lauses esimesena, sellele järgneb täiendav informatsioon *mõlema suuna alusel loodud küsimustikest*, mis allub substantiivile *ülevaate*, mis tekitabki ristumise. Ristumine jääks ära, kui lause oleks *Mõlema suuna alusel loodud küsimustikest ülevaate saamine loob eeldused...*



Joonis 20. Näitelause teadusklassist (EKSP: tea\_eesti\_arst\_2004.inforem)

### 3.2.7. Lahkasendis fraasid

Puudepangas esineb selliseid lauseid, kus (pro)noomeni juurde kuuluvad apositsioonid ei asu tema kõrval, põhjustades ristumise. Lindström (ilmumas) nimetab sellist nähtust postreemaks, kus lahkasendis fraasi ülesandeks on täpsustada pronoomenit. Kokku esines puudepangas 15 lahkasendis fraasist põhjustatud kaarte ristumist, st 100 lause kohta esines mitte-projektiivsust 0.10 korral. Joonisel 21 on pronoomen *nad*, mida laiendavad *need kontrollid*, mis asuvad pronoomenist eemal lause lõpus, mille tõttu tekib ristumine.



Joonis 21. Näitelause ilukirjandusklassist (EKSP: ilu\_sauter\_10000.inforem)

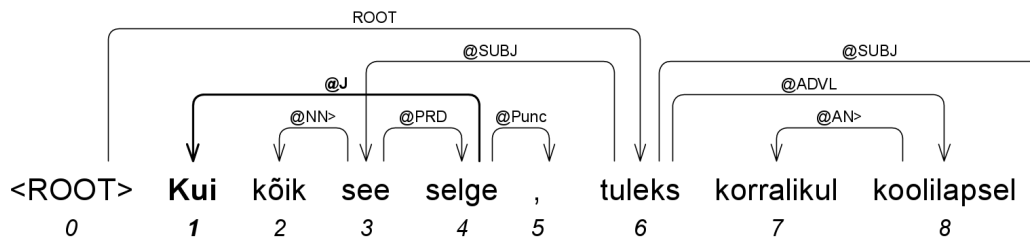
### 3.3. Ellips

Elliptiline ehk väljajätteline lause on lause, kus mingi element jäetakse lausest välja. Tihti peale on selleks kas subjekt või predikaat. Ellips jaguneb kaheks: lühendusellipsiks ja kontekstiellipsiks. Lühendusellipsi puhul piisab lünga täitmiseks ümbrusest ning keeleteadmistest, kontekstiellipsi puhul leiab lüngale lähevormi kontekstist. (EKG II: 222–223)

Kokku esines puudepangas 75 ellipsi põhjustatud kaarte ristumist, st 100 lause kohta esines mitte-projektiivsust 0.48 korral. Puudepangas tekitavad ristumisi peamiselt kontekstiellipsid, aga esineb ka lühendusellipsi poolt põhjustatud ristumisi. Lühendusellipsi puhul jäetakse välja liigne element, nt isikulised asesõnad, *olema* öeldisena või öeldise osana, kõrvallausekorrelaat eelnevas pealauses (EKG II: 223).

Joonisel 22 on välja jäetud verb *olema*, tegelik osalause peaks olema *Kui kõik see on selge. Olema* ellipsit esineb kõnekeeles, kõnekäändudes, vanasõnades, kinnisväljendeis

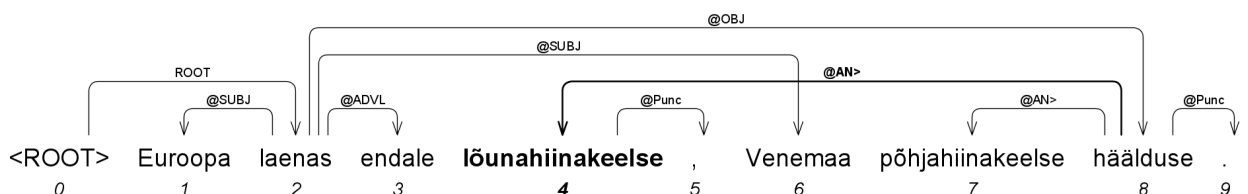
jne (EKG II: 223). Verbi puudumine tingib kaks probleemi, miks kaared ristuvad: 1) verbivormile *tuleks* peaks alluma osalause verbivorm *olema*, kuid selle puududes allub talle *see* 2) konjunktsioon *kui* peaks alluma verbivormile *olema*, kuid selle puududes allub ta predikatiivile *selge*.



Joonis 22. Näitelause ajakirjandusklassist (EKSP: aja\_EPL\_2006\_12\_16.inforem)

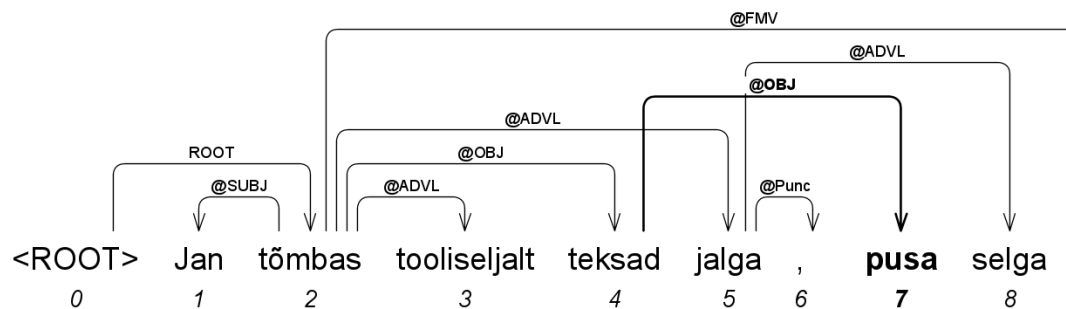
Kontekstiellips on nullasendus, mis asendab lähikontekstis asuvat sõnavormi. Rindlausetes jäetakse tihti välja korduv predikaat või subjekt. Ka ei korrata rinnastuse puhul sarnast laiendit või põhja. (EKG II: 223–225)

Joonisel 23 sõnastatakse moodustaja lõppkomponent ühe korra. Lõppkomponendiks on *häälduse* ja talle allub eestäiend *põhjahiinakeelse*, millele omakorda allub *lõunahiinakeelse*, kuna sellel puudub moodustaja lõppkomponent.



Joonis 23. Näitelause ilukirjandusklassist (EKSP: ilu\_remsu\_siiditeel.inforem)

Joonisel 24 on välja jäetud korduv verbivorm *tõmbas*. Kui muidu alluks verbivormile *tõmbas* objekt *pusa* ja adverbiaal *selga*, siis predikaadi puududes alluvad nad oma samaväärsetele lauseliikmetele, mis põhjustavadki ristumise.



Joonis 24. Näitelause ilukirjandusklassist (EKSP: ilu\_vilep\_lohe.inforem)

### 3.4. Mittelauseliikmelised lauseelemendid

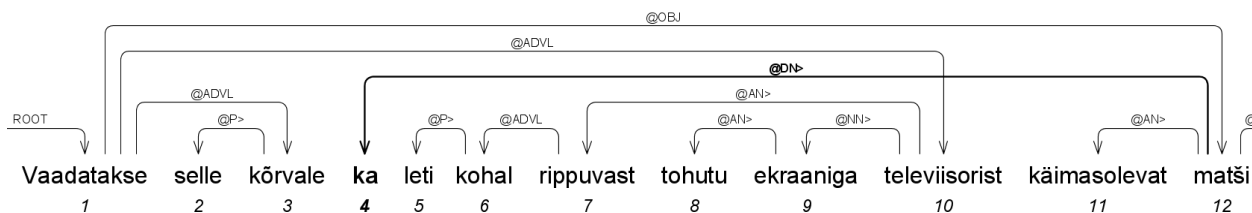
Mittelauseliikmelised lauseelemendid ei laienda ühtegi teist lauseliiget, kuid sellegipoolest modaliseerivad mingit situatsiooni. Mittelauseliikmelisteks elementideks on üldlaiendid, sidendid ja kiilud. (EKG II: 99) Puudepangas põhjustavad ristumisi just üldlaiendid ja kiilud. Kokku esines 27 üldlaiendite põhjustatud kaarte ristumist, st 100 lause kohta esines mitte-projektiivsust 0.11 korral.

#### 3.4.1. Üldlaiend

Üldlaiend on selline lause osa, mille ülesandeks on lauset kuidagi muuta (rõhutamise vmt). Üldlaienditeks on partiklid, vahel ka muudesse sõnaliikidesse kuuluvad sõnad ja nende ühendid. Üldlaiendid on nt *küllap*, *enam*, *ainult*, *pigem*, *ehk*, *küll*, *juba* jne. (EKG II: 99–102) Kokku esines puudepangas 11 üldlaiendi põhjustatud kaarte ristumist, st 100 lause kohta esines mitte-projektiivsust 0.07 korral.

Puudepangas toimub üldlaiendite märgendamine nii, et leitakse, millise sõnaga üldlaiend sisuliselt kõige enam kokku kuulub ning määratakse see tema ülemuseks. Sellest tulenevalt võib üldlaiend olla kas adverb või atribuut. (Märgendusjuhend) Joonisel 25 on üldlaiend *ka*, mis on märgendatud substantiivi *matši* külge. Üldlaiend ristub verbivormi *Vaadatakse* külge kuuluva laiendiga.



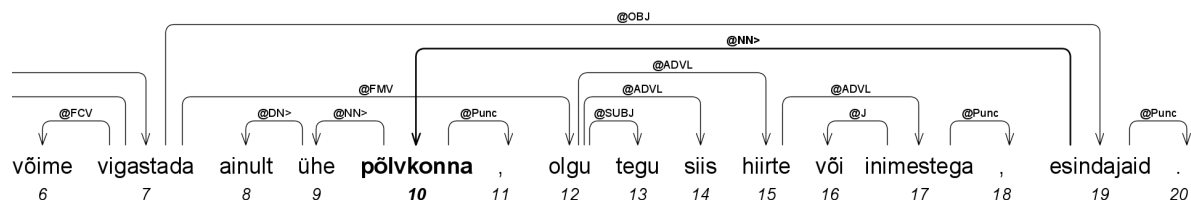


Joonis 25. Näitelause ajakirjandusklassist (EKSP: aja\_EPL\_2006\_12\_06.inforem)

### 3.4.2. Kiil

Kiillause on lause, mis asub teise lause sees, kuid ei mõjuta seda grammatiliselt. Kiillausega antakse edasi mingeid kommentaare, lisainformatsiooni või kõneleja suhtumist. Kiillause eraldatakse komade, mõttekriipsude või sulgude abil. (EKG II: 102–103) Kokku esines puudepangas 6 kiilu põhjustatud kaarte ristumist, st 100 lause kohta esines mitte-projektiivsust 0.04 korral.

Puudepangas riputatakse kiilud osalause peaverbi külge. Ristumised on peamiselt põhjustatud kiilu asetusest lauses. Tihtipeale on ta paigutatud fraasi keskele, mille tõttu tekibki ristumine. Joonisel 26 on kiiluks *olgu tegu siis hiirte või inimestega*, mis on asetatud fraasi *ainult ühe põlvkonna esindajaid* keskele, millest tekib ka ristumine.



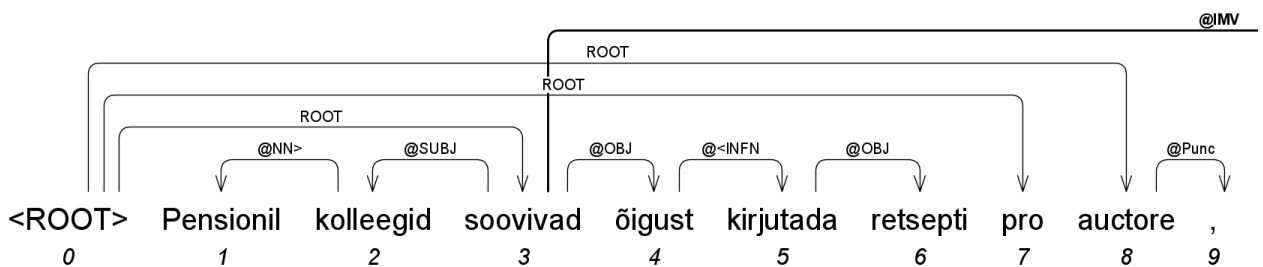
Joonis 26. Näitelause populaarteadusklassist (EKSP: aja\_horisont\_2000\_esimesed\_20000.inforem)

### 3.5. Vead

Puudepangas esineb grupp lauseid, mis on probleemsed lisaks mitte-projektiivusele. See tähendab, et leidub lauseid, kus on mitu juurtippu, mis sisaldavad sulgudes elemente, mis on valesti märgendatud või on juba algtekstis vigased. Kokku esines puudepangas 254 vigadest põhjustatud kaarte ristumist, st 100 lause kohta esines mitte-projektiivsust 1.64 korral.

### 3.5.1. Mitu juurt

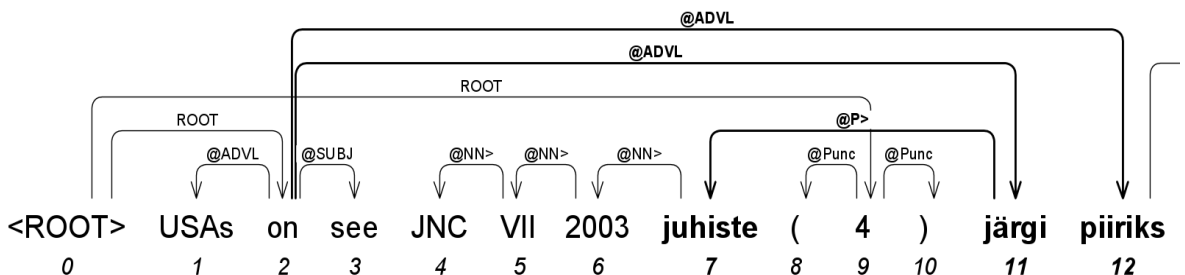
Mitme juure poolt põhjustatud ristuvad kaared tulenevad märgendamisjuhendist. See tähendab, et teatud elemendid, millel puudub keeleline seos, pannakse juurtipu külge (nt võõrkeelsed väljendid *cum laude* jne). Siia kuuluvad ka need vead, mis on tekkinud ConLL-formaati teisendamise tõttu. Kokku esines puudepangas 22 mitme juure põhjustatud kaarte ristumist, st 100 lause kohta esines mitte-projektiivsust 0.14 korral. Joonisel 27 on juure külge pandud *pro auctore*, mis ristub osalauseid ühendava kaarega.



Joonis 27. Näitelause teadusklassist (EKSP: tea\_eesti\_arst\_2004.inforem)

### 3.5.2. Sulgudes element

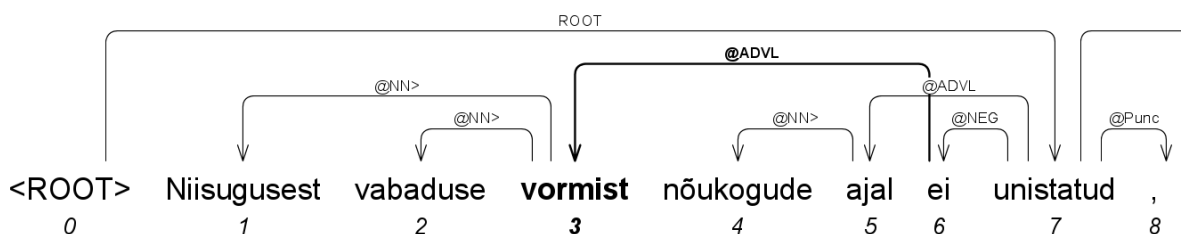
Mitme juure n-ö alagrupiks on sulgudes elemendi märgendamisel tekkinud mitu juurt. Märgendusjuhendist tulenevalt märgendati sulgudes elemente ühel ajahetkel nii, et pandi need kõik juure külge ning teisel hetkel oma loogilise ülemuse külge. Kõik sulgudes elementidega seotud ristumised on pandud kokku siia gruppi. Kokku esines puudepangas 100 sulgudes elemendi põhjustatud kaarte ristumist, st 100 lause kohta esines mitte-projektiivsust 0.65 korral. Joonisel 28 on sulgudes element (4) pandud juure külge, põhjustades ristumise öeldise juurde kuuluvate laienditega.



Joonis 28. Näitelause teadusklassist (EKSP: tea\_eesti\_arst\_2004.inforem)

### 3.5.3. Vale märgendamine

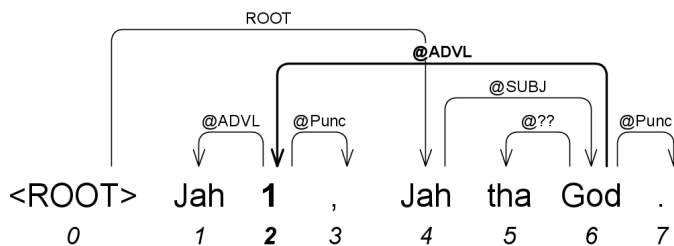
Rohkesti leidub puudepangas lauseid, kus laused on kas valesti märgendatud või märgendamise õigsus on kaheldav. Samuti leidub ka siin lauseid, mis olid ilmselt ConLL-formaati teisendamise tõttu vigaseks muutunud. Kokku esines puudepangas 124 valest märgendamisest põhjustatud kaarte ristumist, st 100 lause kohta esines mitte-projektiivsust 0.80 korral. Joonisel 29 on verbivormiks *unistatud*, millele kuulub *ei*. Täiendid *Niisugusest vabaduse vormist* peaksid kuuluma verbivormile *unistatud*, aga on märgendatud hoopiski *ei* külge. Sellest tuleneb ka ristumine.



Joonis 29. Näitelause ajakirjandusklassist (EKSP: aja\_ee\_2001\_10.tasak.inforem)

### 3.5.4. Vigased laused

Puudepangas esineb selliseid vigaseid lauseid, kus on kas komavead või esinevad mingid elemendid, mida lauses olema ei peaks. Kokku esines puudepangas 8 vigasest lausest põhjustatud kaarte ristumist, st 100 lause kohta esines mitte-projektiivsust 0.05 korral. Joonisel 30 esineb element 1, mis sinna ei kuulu, puudub öeldis.



Joonis 30. Näitelause ilukirjandusklassist (EKSP: ilu\_kanep.tasak.inforem)

## Kokkuvõte

Käesolevas bakalaureusetöös analüüsiti mitte-projektiivsust põhjustavaid süntaktilisi struktuure eesti keele sõltuvuspuude pangas. Leiti, et keskseks probleemiks on mitte-neutraalne sõnajärg kombinatsioonis puudepanga märgendussüsteemiga. Mitte-neutraalne sõnajärg on osadel juhtudel põhjustatud rõhutamisest. Kuna eesti keel on vaba sõnajärgiga, siis on mitte-projektiivsuse olemasolu mõnedes lausetes paratamatus. Sellegipoolest ei esine mitte-projektiivsust eesti keele sõltuvuspuude pangas palju, 100 lausest 4.8 lauses esineb üks või mitu kaarte ristumist.

Esimeses peatükis kirjeldati sõltuvussüntaktilist lähenemist keeleteooriates ja arvutilingvistikas. See sisaldas endas sõltuvusteooria põhimõistete lahtiseletamist, sõltuvuspuu ja fraasistruktuuripuu võrdlust, projektiivsuse lahtimõtestamist ning ülevaadet eesti keele sõltuvuspuude pangast ja seal esinevatest problemaatilistest keelenditest. Teises peatükis kirjeldati materjali ja töö käiku. Kolmas peatükk kirjeldas mitte-projektiivsuse põhjustajaid eesti keele sõltuvuspuude pangas.

Kõige problemaatilisemaks grupiks, mis mitte-projektiivsust põhjustavad, on verbiahelad. Neid märgendatakse hetkel nii, nagu EKG II kirjeldab. See tähendab, et ühte tüüpi verbiahelad analüüsitakse mitmesõnalise öeldisena ning teist tüüpi verbiahelad analüüsitakse nii, et on öeldisverb ja talle kuuluvad laiendid. Kui Universal Dependencies märgendusreeglite järgi verbiahelad analüüsitaks “põõsasteks”, siis kaared üldjuhul ei ristuks (vt viide UD). Joonisel 7 olev verbiahel *pole armunud olnud* oleks analüüsitud nii, et ülemuseks on *armunud* ja talle alluksid nii *pole* kui ka *olnud*. Joonisel 8 olev verbiahel *poleks tohtinud teha* oleks analüüsitud nii, et ülemuseks on *teha* ja talle alluksid nii *poleks* kui ka *tohtinud*. Mõlemal juhul ristumised kaoks.

Teise suure grupi moodustavad sõnajärgiga seotud ristumised. Siia alla kuuluvad kõrvallauseid, kvantor- ja kaassõnafraasid, võrdluskonstruksioonid jne. Mõnede konstruktsioonide puhul saaks üsnagi kergelt ristumised ära kaotada. EKG II järgi on

kvantorfraasi ja kaassõnafaasi põhjaks vastavalt kvantor või kaassõna ning neile allub substantiiv. Kui aga oleks ülemuseks substantiiv, siis ristumised kaoksid üldjuhul ära. Sellist märgendusviisi toetab ka Universal Dependencies (UD). Lisaks on problemaatilised veel elliptilised ja üldlaiendit sisaldavad laused. Viimase grupi moodustavad vigased laused, mis on põhjustatud valest märgendamisest, vigastest lausetest või märgendusjuhendi puudulikkusest.

Töö eesmärgiks oli välja tuua, mis põhjustab mitte-projektiivust lausetes eesti keele sõltuvuspuude pangas. Tulevikus tasub kindlasti uurida, kuidas siin käsitletud mitte-projektiivsuse põhjustajaid analüüsida nii, et kaared enam ei ristuks.

## Kirjandus

**EKG II = Erelt, Mati, Reet Kasik, Helle Metslang, Henno Rajandi, Kristiina Ross, Henn Saari, Kaja Tael, Silvi Vare 1993.** Eesti keele grammatika II. Süntaks. Tallinn: Eesti Teaduste Akadeemia Keele ja Kirjanduse Instituut.

**Jurafsky, Daniel, James H. Martin 2009.** Speech and Language Processing: An introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. 2nd edition. Upper Saddle River (N.J.): Pearson/Prentice Hall, 438.

**Liin, Krista 2014.** Puudepangad. Arvutilingvistika alused konspekt. Tartu Ülikool: eesti ja üldkeeleteaduse instituut. Kasutatud 20.05.2016

**Lindström, Liina (ilmumas).** Infostruktuur ja lause sõnajärg. – Eesti keele süntaks. (Eesti keele varamu 3.) Tartu: Tartu Ülikooli kirjastus.

**McDonald et al 2013 = McDonald, Ryan, Joakim Nivre, Yvonne Quirmbach-Brundage, Yoav Goldberg, Dipanjan Das, Kuzman Ganchev, Keith Hall, Slav Petrov, Hao Zhang, Oscar Täckström, Claudia Bedini, Núria Bertomeu Castelló, and Jungmee Lee 2013.** Universal Dependency Annotation for Multilingual Parsing. – Proceedings of ACL, 92–97. <http://www.aclweb.org/anthology/P13-2017>. Vaadatud 25. mai 2016.

**Melčuk, Igor A. 1988.** Dependency Syntax: Theory and Practice. State University of New York Press.

**Muischnek, Kadri 2016.** Suuline viide. Tartu: 20. mai.

**Muischnek jt 2014a = Muischnek, Kadri, Kaili Müürisep, Tiina Puolakainen 2014.**

Dependency Parsing of Estonian: Statistical and Rule-based Approaches. – Human Language Technologies – The Baltic Perspective. Eds. A. Utko et al., 111–118. <http://ebooks.iospress.nl/publication/38013>. Vaadatud 24.05.2016.

**Muischnek jt 2014b = Muischnek, Kadri, Kaili Müürisep, Tiina Puolakainen, Eleri Aedmaa, Riin Kirt, Dage Särg 2014.** Estonian Dependency Treebank and its annotation scheme. – Proceedings of the Thirteenth International Workshop on Treebanks and Linguistic Theories (TLT13). Tübingen, Saksamaa: University of Tübingen, 285–297. <http://tlt13.sfs.uni-tuebingen.de/tlt13-proceedings.pdf>. Vaadatud 15. mai 2016.

**Müürisep, Kaili 2000.** Eesti keele arvutigrammatika: süntaks. Dissertationes Mathematicae Universitatis Tartuensis 22. Tartu

**Nilsson, Jens, Joakim Nivre 2008.** MaltEval: An Evaluation and Visualization Tool for Dependency Parsing. – Proceedings of LREC, 161-166. [http://lrec.elra.info/proceedings/lrec2008/pdf/52\\_paper.pdf](http://lrec.elra.info/proceedings/lrec2008/pdf/52_paper.pdf). Vaadatud 20. mai 2016.

**Nivre, Joakim 2005.** Dependency grammar and dependency parsing. – MSI report 05133, School of Mathematics and Systems Engineering, Växjö University. <http://stp.lingfil.uu.se/~nivre/docs/05133.pdf>. Vaadatud 15. mai 2016.

### **Internetimaterjalid:**

**EKSP = Eesti keele sõltuvuspuude pank;** <https://www.keeletehnoloogia.ee/et/ekt-projektid/vahendid-teksti-mitmekihiliseks-margendamiseks-rakendatuna-koondkorpusele/soltuvussuntaktiliselt-analuusitud-korpus>. Vaadatud 20. mai 2016.

**Märgendusjuhend** = <https://www.keeletehnoloogia.ee/et/ekt-projektid/vahendid-teksti-mitmekihiliseks-margendamiseks-rakendatuna-koondkorpusele/soltuvussuntaktiliselt->

analüüsitud-korpus. Vaadatud 20. mai 2016.

**UD = Universal Dependencies märgendusreeglid;**

<http://universaldependencies.org/u/overview/syntax.html>. Vaadatud 24. mai 2016.



## **Non-projective sentences in the Estonian Dependency Treebank**

### **Summary**

Treebank or syntactically annotated corpus plays an important role in natural language processing. It enables the development of machine translation systems and the testing of both language software and theories (Liin 2014). The Estonian Dependency Treebank consists of a morphological, surface-syntactic and a dependency layer (Muischnek *et al* 2014b).

This work focuses particularly on dependency relations. These dependency relations are described by arcs which denote which words are dependent on which other words. Theoretical approaches to language assume that these arcs do not cross (this property is called projectivity), but admit that crossing of dependency arcs (non-projectivity) is unavoidable for the representation of some linguistic structures (Nivre 2005). Computational linguistics tries to avoid crossing of arcs for practical purposes.

The goal of this Bachelor's thesis is to describe the syntactical structures which cause non-projectivity in the Estonian Dependency Treebank. The first chapter gives an overview of the dependency-syntactical approach to language theory and computational linguistics. This chapter contains an explanation of the fundamental terms of dependency theory, a comparison of dependency trees and phrase structure trees, and an overview of the Estonian Dependency Treebank and the problematic linguistic forms that it contains. The second chapter describes the materials used and the work process. The third chapter describes syntactical structures which cause non-projectivity. The fourth chapter gives an overview of the results.

**Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks**

Mina

Liisi Torga

(*autori nimi*)

(sünnikuupäev:

06.12.1992

)

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose

“Mitte-projektiivsed laused eesti keele sõltuvuspuude pangas”

(*lõputöö pealkiri*)

mille juhendaja on

Kadri Muischnek

(*juhendaja nimi*)

- 1.1. reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil, sealhulgas digitaalarhiivi DSpace-is lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;
- 1.2. üldsusele kättesaadavaks tegemiseks Tartu Ülikooli veebikeskkonna kaudu, sealhulgas digitaalarhiivi DSpace'i kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.
2. olen teadlik, et punktis 1 nimetatud õigused jäävad alles ka autorile.
3. kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

Tartus, 30. mai 2016 (*kuupäev*)